KOUTILYA PANDE

New Brunswick, NJ |+1 (929) 3785340 | koutilyapande26@gmail.com | LinkedIn

EDUCATION

RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY *Sept'22 – Dec'23*
*Master of Information Technology and Analytics, Major :- **Data Science***
SHRI GOVINDRAM SEKSARIA INSTITUTE OF TECHNOLOGY AND SCIENCE, INDIA *Aug'18 – May'22*
Bachelor of Technology (B. Tech) : Electronics and Telecommunication Engineering

SKILLS

Languages: C, C++,R, Python (Pandas, NumPy, Matplotlib, TensorFlow, Scikit Learn, Keras), SQL
Functional: Data Modelling, Statistical Analysis, Data Visualization, Transformers, Huggingface NLP, Gen AI, Langchain, RAG
Database: MS SQL, My SQL, Snowflake, MongoDB,SSRS, ADF, S3 Buckets
Tools and Platforms: MS Excel, Jupyter Notebook, AWS, Azure AI, GitHub, Tableau, R Studio, PowerBI, Streamlit, LLMs

EXPERIENCE

Data Scientist *Mar'24 – Present*
Keelworks Foundation Oak Harbor,USA
- Led data collection and data preprocessing team, ensuring the data quality and integrity for models
- Implemented **sentiment analysis** using **BERT** and **NER** using **spaCy** with fellow machine learning engineers for improving customer feedback categorization and enhanced information extraction efficiency by 40%
- Optimized homepage layout by conducting **A/B testing** and **statistical analysis**, resulting in a **28%** increase in **CTR**
- Tracked and analyzed core business KPIs using **PowerBI** and **Excel**, highlighting insights on user behavior

Data Scientist *Dec'21 – Jul'22*
Korangle Indore,India
- Automated **data sampling** using **Python**, reducing processing time by **40%** while conducting rigorous **statistical analyses** on 50,000+ entries for data validation, ensuring **data integrity** for product development and marketing strategies
- Developed a customer segmentation model using **K-Means clustering** to find new set of customers
- Analyzed **competitive market data** and **customer queries** using NLP increasing market penetration by **17%**
- Assisted with customer journey mapping, Led a **Random Forest customer satisfaction** model from scratch, synthesizing data from customer surveys, resulting in **12%** fewer customer complaints

Data Analyst *May'21 – July'21*
Udemy Indore,India
- Designed queries, stored procedures, triggers, cursors, tables, and views in **SQL Server** to provide structured data for generating comprehensive **SSRS** analytical reports
- Optimized **ETL** by leveraging **SQL and Python**, **18% reduction** in data preparation time, saving **8hr/week** for analytic team
- Devised **classification models**(Logistic Regression, Decision Tree, Random Forest)to predict loan approval outcomes
- Successfully addressed **class imbalance** in loan approval prediction model by conducting **extensive model evaluation** (precision, recall, F1 score, AUC-ROC)and fine-tuning to achieve accuracy of **84.06%**

PROJECTS

CoverMe.ai |(Private Repositpry) | *(MERN stack, OpenAI, Cursor, v0)*
- SaaS Chrome extension which allows users to generate tailored cover letters while utilizing efficient APIs and optmized prompts with a complete website and profile page.

PDF ChatGenie | GitHub **|** *(RAG, Streamlit, LangGraph, LangChain)*
- Developed a multi-PDF **conversational chatbot** application with AI tools for providing contextual aware answers through RAG and provide human-like chat interaction with system message

Naruto Series Analyzer - NLP | Github **|** *(Transformers, HuggingFace, LORA, Gradio, Llama)*
- Built a Gradio web application using SpaCy for NER and **BERT-based zero-shot classifiers** to analyze character interactions and themes from scraped datasets using Scrapy
- Embedded custom jujutusu classifiers using **DistilBERT** and a Naruto chatbot using Transformers and Huggingface, **fine-tuning** Llama 3.2 using **LORA** showcasing an end-to-end NLP pipeline with real-time user interaction.

Dodging The Mishap - **Pandemic Predictor Web Application |** GitHub **|** *(Time-Series, LSTM, Django, YFinance)*
- Designed a **Django** web application powered by an **LSTM** machine learning model, performed **time series analysis** on stock prices based on real time financial data from YFinance

MoodSync: Speech-Driven Spotify Song Recommendations | GitHub **|** *(Python, Tensorflow, sklearn)*
- Authored a **Recommendation System** deploying **Multi-Layer Perceptron Classifier** from scrath to recognize emotions through speech **Librosa** and recommended Spotify playlists based on identified emotion

BCG Data Science Job Simulation: Forage | GitHub **|** *(Python, EDA, Feature Engineering, Hypothesis Testing)*
- Completed **customer churn prediction analysis simulation** on SME customers for PowerCo. with **87%** accuracy

AI Agent
- Financial Agent with Web Search News Tool | GitHub **|** *(phidata, FASTAPI)*
- Essay Grading Agent | GitHub **|** *(LangGraph, OpenAI)*
- Blog Writing Agent | GitHub **|** *(OpenAI SWARM, Tavily)*